Linearly Representable Entropy Vectors and their Relation to Network Coding Solutions

Asaf Cohen, Michelle Effros, Salman Avestimehr and Ralf Koetter

***Abstract*—In this work, we address the following question: "When can we guarantee the optimality of linear coding at all internal nodes of a network?" While sufficient conditions for linear coding throughout the network are known, it is not clear whether relaxing the linearity constraints at the terminal nodes can result in simpler operations at the internal nodes.**

We present a novel method to analyze the space of network solutions using the constraints resulting from the network topology, and we identify sufficient conditions for an optimal linear solution at all internal nodes. These conditions are also sufficient to characterize the rate region only in terms of Shannon information inequalities.

I. Introduction

Optimal coding strategies for networks are well understood only for some demands. For example, it is well known that linear coding suffices for multicast [1], [2]. For general demands, mainly negative results are available, e.g. cases where linear codes are suboptimal [3].

In this work, we take advantage of the functional constraints within the network in order to limit the space of possible network solutions, and we derive sufficient conditions under which the resulting space adheres a linear representation at all internal nodes. Network codes which employ non-linear operations at the terminal nodes together with linear network coding at the internal nodes are useful, for example, in non-multicast networks with dependent sources, such as networks with side information [4].

While similar constraints where used in the LP bound [5] and the Ingleton-LP bound [6], [7], the focus in this work is not on deriving tight outer bounds, but rather on when can a linear representation be found for a given network and what are the applications such a representation has on the codes that can be used.

Furthermore, an important outcome of this representation is the ability to test whether Shannon-type inequalities suffice in order to derive the rate region of a network. Non-Shannon inequalities might not be linear (e.g. [8]), cannot be expressed in a canonical form to facilitate the analysis, and, most importantly, may or may not be known to us in their entirety. (For a comprehensive tutorial see [9].) For these reasons, even in the absence of a complete characterization of the rate region, it is desirable to diminish the need for non-Shannon inequalities.

Any network coding solution can be viewed as an optimization over the entropic region subject to constraints imposed by the network topology and the receiver demands. Therefore, the focus has been on characterizing the entropic region. This characterization turns out to be quite difficult, and remains an open problem for more than three variables. Our contribution is to give sufficient conditions under which the Shannon inequalities together with the network constraints result in an outer bound which is completely within the entropic region. In this case, there is no need to consider non-Shannon inequalities. Moreover, properties which apply to the outer bound, immediately apply to any solution for the network. As a result, we are able to prove interesting properties for rate regions on networks with non-multicast demands, a class of networks which is generally unsolved.

The rest of the paper is organized as follows. Section II introduces definitions and previous results. Section III contains the main result. Section IV includes two examples.

II. Preliminaries

A. Networks and Codes

A network is defined as a directed graph $(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of vertices (nodes) and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges (links). For each edge $e = (a, b) \in \mathcal{E}$, we use $o(e) = a$ and $d(e) = b$ to denote the origin and destination vertices, respectively, of edge e. Associated with each edge $e \in \mathcal{E}$ is a capacity $c(e) \geq 0$. We assume acyclic graphs and noise-free links.

Let $\{(X_{1,i}, \ldots, X_{K,i})\}_{i=1}^{\infty}$ be a sequence of independent and identically distributed K-tuples of discrete random variables with alphabet $\mathcal{X}_1 \times \cdots \times \mathcal{X}_K$. We refer to X_j as the j-th source in the network. The sources are statistically independent with some known distribution $\Pi_{i=1}^{K} p_i(x_i)$. Let $\mathcal{K}$ be the power set of $\{1, \ldots, K\}$ and denote by $S : \mathcal{V} \mapsto \mathcal{K}$ the mapping defining the sources available at each node. Analogously, denote by $D : \mathcal{V} \mapsto \mathcal{K}$ the mapping defining the sources demanded at each node. We assume that nodes for which $S(v) \neq \emptyset$, i.e. *source nodes*, have no incoming edges and that nodes for which $D(v) \neq \emptyset$, i.e. *terminal nodes* have no outgoing edges. A network with graph $(\mathcal{V}, \mathcal{E})$, sources $\{X_i\}_{i=1}^{K}$ and mappings S and D is denoted by $(\mathcal{V}, \mathcal{E}, \Pi_{i=1}^{K} p_i(x_i), S, D)$.

[0]Asaf Cohen was with the California Institute of Technology. He is now with the Department of Communication System Engineering, Ben-Gurion University of the Negev, Israel (coasaf@bgu.ac.il).
Michelle Effros is with the Department of Electrical Engineering, California Institute of Technology, Pasadena, CA 91125 (effros@caltech.edu).
Salman Avestimehr is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 (avestimehr@ece.cornell.edu).
Ralf Koetter is with the Institute for Comm. Engineering, Technische Universität München, D-80290 München, Germany (ralf.koetter@tum.de).
This work is partially supported by DARPA Grant W911NF-07-I0029, Lee Center for Advanced Networking at Caltech and the Center for Mathematics of Information at Caltech.

Definition 1. For any vector of rates $(R_e)_{e\in\mathcal{E}}$, a $((2^{nR_e})_{e\in\mathcal{E}}, n)$ *source code* is the following set of mappings

$$g_e^n : \Pi_{i\in S(o(e))}\mathcal{X}_i^n \mapsto \{1,\ldots,2^{nR_e}\} \quad e\in\mathcal{E}, S(o(e))\neq\emptyset$$
$$g_e^n : \Pi_{e':d(e')=o(e)}\{1,\ldots,2^{nR_{e'}}\} \mapsto \{1,\ldots,2^{nR_e}\} \quad e\in\mathcal{E}, S(o(e))=\emptyset$$
$$h_t^n : \Pi_{e:d(e)=t}\{1,\ldots,2^{nR_e}\} \mapsto \Pi_{i\in D(t)}\mathcal{X}_i^n \quad D(t)\neq\emptyset.$$

For each terminal node $t\in\mathcal{V}$ we use $\hat{X}_{i,t}^n$ to denote the reproduction of X_i^n found by decoder h_t^n. We are interested in the set of possible values $(c(e))_{e\in\mathcal{E}}$ for which the sources can be reproduced at the terminals with a negligible error probability. Precisely, we require that for any $\epsilon>0$ there exists a sufficiently large n and a $((2^{nR_e})_{e\in\mathcal{E}}, n)$ code with $R_e\leq c(e)$ for all $e\in\mathcal{E}$, such that $\Pr(\hat{X}_{i,t}^n = X_i^n)\geq 1-\epsilon$ for all terminal nodes t and demands $i\in D(t)$. We call the closure of this set of rate vectors the *set of achievable rates*, which we denote by $\mathcal{R}(\mathcal{V},\mathcal{E},\Pi_{i=1}^K p_i(x_i), S, D)$.

B. Polymatroids, Entropic and Pseudo-Entropic Vectors

Let N be an index set of size n and $\mathcal{N}$ be the power set of N. A function $g:\mathcal{N}\mapsto\mathbb{R}$ defines a *polymatroid* (N,g) with a ground set N and rank function g if it satisfies the following three conditions [10]:

$$g(\emptyset) = 0, \tag{1}$$
$$g(I) \leq g(J) \text{ for } I\subseteq J\subseteq N, \tag{2}$$
$$g(I)+g(J) \geq g(I\cup J)+g(I\cap J) \text{ for } I,J\subseteq N. \tag{3}$$

For any polymatroid g with ground set N, we can represent g by the vector $(g(I): I\subseteq N)\in\mathbb{R}^{2^n-1}$ defined on the ordered, non-empty subsets of N. We denote the set of all polymatroids with a ground set of size n by Γ_n. Thus $\mathbf{w}\in\Gamma_n$ if and only if w_I and w_J satisfy equations (1)–(3) for all $I,J\subseteq N$, where w_I is the value of $\mathbf{w}$ at the entry corresponding to the subset I.

Let the ground set N be a set of discrete random variables. For any $A\subseteq N$, let $H(A)$ be the joint entropy function. Throughout, entropies are measured in bits, thus $\log()$ denotes the base-2 logarithm. An entropy vector $\mathbf{w}$ is a (2^n-1)-dimensional vector whose entries are the joint entropies of all non-empty subsets of N. It is well-know that the entropy function is a polymatroid over this ground set N. In fact, the polymatroid axioms are equivalent to the basic information inequalities [9]. However, the opposite is not necessarily true. That is, there exists points $\mathbf{w}\in\Gamma_n$ $(n>3)$ for which there is no set of n discrete random variables whose joint entropies equal $\mathbf{w}$. Following [6], we denote by Γ_n^* the set of all $\mathbf{w}\in\Gamma_n$ for which there exists at least one random vector whose joint entropies equal $\mathbf{w}$. A $\mathbf{w}\in\Gamma_n$ is called *pseudo-entropic*, while if this $\mathbf{w}$ is also in Γ_n^* it is called *entropic*.

Denote by $\bar{\Gamma}_n^*$ the convex closure of Γ_n^*. Then $\bar{\Gamma}_n^*=\Gamma_n$ for $n\leq 3$ but $\bar{\Gamma}_n^*\neq\Gamma_n$ for $n>3$ [9].

C. Polyhedral Cones

For an integer $d>0$, a *polyhedral cone* C is the set of all vectors $\mathbf{x}\in\mathbb{R}^d$ satisfying $A\mathbf{x}\geq 0$ for some $A\in\mathbb{R}^{m\times d}$. That is

$$C=\{\mathbf{x}\in\mathbb{R}^d : A\mathbf{x}\geq 0\}.$$

Let $\mathbf{u}_1,\ldots,\mathbf{u}_p$ be a set of vectors in $\mathbb{R}^d$. Denote by $\text{NonNeg}(\mathbf{u}_1,\ldots,\mathbf{u}_p)$ the non-negative, or *conic*, hull of $\mathbf{u}_1,\ldots,\mathbf{u}_p$, namely,

$$\text{NonNeg}(\mathbf{u}_1,\ldots,\mathbf{u}_p) = \{\lambda_1\mathbf{u}_1+\ldots+\lambda_p\mathbf{u}_p : \lambda_i\in\mathbb{R}, \lambda_i\geq 0 \quad i=1,\ldots,p\}. \tag{4}$$

Theorem 1 (Minkowski, e.g. [11])**.** *For any polyhedral cone C in $\mathbb{R}^d$ there exist vectors $\mathbf{u}_1,\ldots,\mathbf{u}_p$ such that*

$$C=\text{NonNeg}(\mathbf{u}_1,\ldots,\mathbf{u}_p).$$

The current literature includes several algorithms (e.g. [11]) to retrieve one representation of a polyhedral cone from the other. In [12], the authors used the representation of a polyhedral cone given in Theorem 1 to show that the polymatroid axioms and Ingleton inequality characterize all inequalities for ranks of up to 4 linear subspaces. In this work, we use this representation to analyze network coding solutions.

D. Finite Alphabet Linear Representation

Let $\mathcal{L}$ be a linear (vector) space over a finite field F. Let $\{L_i\}_{i=1}^n$ be subspaces of $\mathcal{L}$. Let $\mathbf{w}(\{L_i\})$ be the (2^n-1)-dimensional vector whose entries are the ranks of all possible unions of L_i's. That is

$$\mathbf{w}(\{L_i\}) = (\text{rank}(L_1),\ldots,\text{rank}(L_n), \text{rank}(L_1\cup L_2),\ldots,\text{rank}(\cup_{i=1}^n L_i)) \tag{5}$$

Theorem 2 ([12, Theorem 2])**.** *For any finite linear space $\mathcal{L}$ over a finite field F or the real line $\mathbb{R}$, and any set of subspaces $\{L_i\}$, $\mathbf{w}(\{L_i\})$ is entropic.*

Proof: We include here a short description of how to construct a random vector whose entropy vector is proportional to $\mathbf{w}(\{L_i\})$. This construction is similar to that in [12, Theorem 2], but our approach facilitates a more straight-forward description of the corresponding source code, which we derive in the next section.

Let $\mathbf{v}_1,\ldots,\mathbf{v}_k$ be a basis for $\mathcal{L}$. For each $\mathbf{v}_i$, assign a random variable V_i, distributed uniformly on F and independently of all other V_j, $j\neq i$. For each L_i, we write

$$L_i=\text{span}\left(\mathbf{l}_1^i,\ldots,\mathbf{l}_{k_i}^i\right)$$

where $\text{rank}(L_i)=k_i$ and for each $\mathbf{l}_j^i$ we have

$$\mathbf{l}_j^i=\sum_{m=1}^k \beta_m^{i,j}\mathbf{v}_m.$$

The random variable corresponding to L_i is defined by

$$X_i=\left(\sum_{m=1}^k \beta_m^{i,1}V_m,\ldots,\sum_{m=1}^k \beta_m^{i,k_i}V_m\right) \tag{6}$$

over the alphabet F^{k_i}. Since the coefficient vectors $(\{\beta_m^{i,j}\}_{m=1}^k)$ for $j = 1,\ldots,k_i$ are linearly independent, we have $H(X_i) = k_i \log|F|$. Define $X_I = (\{X_i\}_{i\in I})$. It is not hard to show that $H(X_I) = \text{rank}(\cup_{i\in I} L_i)\log|F|$. For example, consider the pair (X_i, X_j). Although it takes values on $F^{k_i+k_j}$, the number of distinct values it can take is $F^{\text{rank}(L_i\cup L_j)}$, all with uniform probability, and thus $H(X_i, X_j) = \text{rank}(L_i \cup L_j)\log|F|$. ∎

The following example is useful for understanding the derivations in the next section.

Example 1. Let $\mathcal{L}$ be a linear space of rank k with subspaces L_1, L_2, L_3. Assume that

$$\text{rank}(L_1 \cup L_2) = \text{rank}(L_1 \cup L_2 \cup L_3).$$

That is, $L_3 \subseteq \text{span}(L_1, L_2)$. Let the corresponding random variables X_i, $i \in \{1,2,3\}$, be defined according to (6). Since each of the vectors $(\{\beta_m^{3,1}\}_{m=1}^k)$ to $(\{\beta_m^{3,k_3}\}_{m=1}^k)$ is a linear combination of the β^1 and β^2 vectors, we have $X_3 = G(X_1, X_2)^T$, where

$$G \in F^{\text{rank}(L_3)} \times F^{(\text{rank}(L_1)+\text{rank}(L_2))}.$$

Definition 2. Let $\mathbf{w}$ be in Γ_n^*. If there exists a linear space $\mathcal{L}$ over a finite field F and n subspaces $\{L_i\}_{i=1}^n$ such that $\mathbf{w} = \alpha\mathbf{w}(\{L_i\})$ for some $\alpha > 0$ we say that $\mathbf{w}$ has a *linear representation over a finite field* F.

Example 2. Consider the following $\mathbf{w} \in \Gamma_3$

$$\mathbf{w} = (1,1,1,2,2,2,2).$$

Let e_1 and e_2 be independent vectors over $GF(2)$. Then $\mathcal{L} = \text{span}(\{e_1, e_2\})$, $L_1 = \text{span}(\{e_1\})$, $L_2 = \text{span}(\{e_2\})$ and $L_3 = \text{span}(\{e_1+e_2\})$ gives $\mathbf{w}(\{L_i\}) = \mathbf{w}$. Thus $\mathbf{w}$ is binary linearly representable and by Theorem 2 it is entropic. It is easy to see that $\mathbf{w}$ is the entropy vector of $(X_1, X_2, X_1 \oplus X_2)$ where X_1 and X_2 are independent uniform random bits.

III. Results

Let $(\mathcal{V}, \mathcal{E}, \Pi_{i=1}^K p_i(x_i), S, D)$ be a given network. Let $\{1,\ldots,K+|\mathcal{E}|\}$ be the ground set of $\Gamma_{K+|\mathcal{E}|}$. We assume the first K entries, $w_1,\ldots,w_K$ represent the K information sources while the following $|\mathcal{E}|$ entries, $w_{K+1},\ldots,w_{K+|\mathcal{E}|}$, represent the $|\mathcal{E}|$ edges in the network. Similar to the definitions in Section II-B, the remaining $2^{K+|\mathcal{E}|}-K-|\mathcal{E}|-1$ labels represent the remaining non-empty subsets of $\{1,\ldots,K+|\mathcal{E}|\}$ in a lexicographic order. That is, w_A, $A \subseteq \{1,\ldots,K+|\mathcal{E}|\}$ is the entry of $\mathbf{w}$ corresponding to the subset A.

Definition 3. Denote by R_{out} the set of all pseudo-entropy vectors $\mathbf{w} \in \Gamma_{K+|\mathcal{E}|}$ for which

$$w_{X_1^K} - \sum_{i=1}^K w_{X_i} = 0 \tag{7}$$

$$w_{e\cup S(o(e))} - w_{S(o(e))} = 0 \quad \forall e : S(o(e)) \neq \emptyset \tag{8}$$

$$w_{e\cup\{e':d(e')=o(e)\}} - w_{\{e':d(e')=o(e)\}} = 0 \quad \forall e : S(o(e)) = \emptyset \tag{9}$$

$$w_{D(t)\cup\{e:d(e)=t\}} - w_{\{e:d(e)=t\}} = 0 \quad \forall t : D(t) \neq \emptyset \tag{10}$$

Equations (7)-(10) require a few remarks. First, note that R_{out} is a subset of $\Gamma_{K+|\mathcal{E}|}$, hence $\mathbf{w} \in R_{out}$ does not necessarily represent an entropic vector. That is, equation (7) should interpreted as follows: in the pseudo-entropy vector $\mathbf{w} \in \Gamma_{K+|\mathcal{E}|}$, the entry corresponding to all sources is equal to the sum of the entries corresponding to the individual sources. Had $\mathbf{w}$ been entropic, equation (7) would mean $H(X_1^K) = \sum_{i=1}^K H(X_i)$. Note that for a network code using a fixed block length n, requiring the independence of $(X_1)^n,\ldots,(X_K)^n$ would result in the same linear constraint on the points in $\Gamma_{K+|\mathcal{E}|}$, namely, equation (7). The rest of the equations have an analogous interpretation. Equation (8) means that the entry corresponding to an outgoing edge of a source node is equal to the entry corresponding to that edge and the sources available at that node. For entropic $\mathbf{w}$, this means that the entropy of the random variable assigned to an edge e for which $S(o(e)) \neq \emptyset$ is zero conditioned on the sources available at $o(e)$. This should be true for any fixed block used, that is, if the sources available at a node are $(X_i)^n,\ldots,(X_j)^n$, then each outgoing edge if a function of these blocks. Equations (9) and (10) have an analogous interpretation. To conclude, equations (7)-(10) define a polyhedral cone in $\Gamma_{K+|\mathcal{E}|}$ which represents the independence, functional and decoding constraints in the network.

For any subset $Q \subseteq \mathbb{R}^d$, denote by $Q|_A$ the projection of Q on the set of coordinates $A \subseteq \{1,\ldots,d\}$. The following lemma gives an outer bound on the rate region of a network. It is very similar to [9, Theorem 15.9] in the usage of the independence, functional and decoding constraints in the network in order to bound the rate region, and is included here for completeness. The main contribution of this work is not the actual outer bound R_{out}, which, in fact, can be replaced by any polyhedral outer bound, but the proof that if the region R_{out} satisfies certain properties (linear representability) then the rate region also has desirable properties. Specifically, if R_{out} is linearly representable, then any point in the rate region can be implemented using simple linear operations in the internal nodes.

Lemma 1. *Let $(\mathcal{V}, \mathcal{E}, \Pi_{i=1}^K p_i(x_i), S, D)$ be a given network. Let R_{out} be defined according to Definition 3. Then $\mathcal{R}(\mathcal{V}, \mathcal{E}, \Pi_{i=1}^K p_i(x_i), S, D) \subseteq R_{out}|_{\{K+1,\ldots,K+|\mathcal{E}|\}}$.*

The inequalities composing R_{out} are linear, hence result in a polyhedral cone representation for R_{out} and Theorem 1 applies directly, giving

$$R_{out} = \text{NonNeg}(\mathbf{w}_1,\ldots,\mathbf{w}_m), \quad \{\mathbf{w}_i\}_{i=1}^m \subseteq \Gamma_{K+|\mathcal{E}|}. \tag{11}$$

The representation given in (11) is useful because it allows us to easily[1] derive R_{out} for specific small networks (using, for example, the software given in [11]).

The following theorem is the main result in this work. It gives a sufficient condition under which it suffices to

[1]Note that although the complexity of computing this representation is exponential in the problem dimension, the constraints (7)-(10) reduce the actual dimension significantly.

use linear coding operations at all internal nodes. Moreover, under this condition, although to date there is no complete characterization of $\bar{\Gamma}^*_{K+|\mathcal{E}|}$, we know that the outer bound on the rate region is completely inside $\bar{\Gamma}^*_{K+|\mathcal{E}|}$. As a result, no non-Shannon inequalities govern the rate region of the network.

Theorem 3. *Let $(\mathcal{V}, \mathcal{E}, \Pi_{i=1}^K p_i(x_i), S, D)$ be a given network. Let R_{out} be defined according to Definition 3 and let $\{\mathbf{w}_i\}_{i=1}^m$ be pseudo-entropy vectors such that $R_{out} = \text{NonNeg}(\mathbf{w}_1, \ldots, \mathbf{w}_m)$. Then, if $\{\mathbf{w}_i\}_{i=1}^m$ are linearly representable over a finite field F, the following is true*

1) *Any point in $\mathcal{R}(\mathcal{V}, \mathcal{E}, \Pi_{i=1}^K p_i(x_i), S, D)$ is achievable with linear coding at all but the source and terminal nodes.*
2) *$R_{out} \subseteq \bar{\Gamma}^*_{K+|\mathcal{E}|}$. Consequently, the rate region of the network is not determined by non-Shannon information inequalities.*

Before we prove Theorem 3, we present the following lemma, which states that any point within the cone generated by linearly representable vectors can be *approximated* by linearly representable vectors.

Lemma 2. *For any $\mathbf{w} \in \text{NonNeg}(\mathbf{w}_1, \ldots, \mathbf{w}_m)$, where $(\mathbf{w}_1, \ldots, \mathbf{w}_m)$ are linearly representable over a finite field F, there exists a sequence of linearly representable vectors $\{\mathbf{r}_\eta\}$ such that $\lim_{\eta\to\infty} \|\mathbf{r}_\eta - \mathbf{w}\|_\infty = 0$.*

Lemma 2 implies that $\mathbf{w} \in \text{NonNeg}(\mathbf{w}_1, \ldots, \mathbf{w}_m)$ is asymptotically entropic, and hence $\text{NonNeg}(\mathbf{w}_1, \ldots, \mathbf{w}_m) \subseteq \bar{\Gamma}^*_l$, where l is the size of the ground set. The proof is not, however, a direct extension of [13, Theorem 1], where the authors prove that $\bar{\Gamma}^*_n$ is a convex cone. The key is that one cannot use *any* auxiliary random variable to perform the convex combination between the random vectors since we require that all $\mathbf{r}_\eta$ are linearly representable. The complete proof appears in [14].

proof sketch (Theorem 3): If $\omega \in \mathcal{R}$, then for any $\epsilon > 0$ there exists a block length n, encoding functions g_e^n such that

$$\frac{1}{n} \log \|g_e^n\| \leq R_e + \epsilon, \tag{12}$$

where R_e is the rate constraint on the edge e, and decoding functions h_t^n such that $\Delta_t \leq \epsilon, \quad t \in T$, where Δ_t is the probability of error in reconstructing the demands at terminal t. For information sources $X_1, \ldots, X_K$, this code defines the random vector

$$((X_1)^n, \ldots, (X_K)^n, \{\hat{g}_e^n((X_1)^n, \ldots, (X_K)^n)\}_{e\in\mathcal{E}}),$$

where $\hat{g}_e^n$ represents the global coding function on edge e. Denote its entropy vector by $\mathbf{h}_{C_n}$. The entropy vector $\mathbf{h}_{C_n}$ satisfies $\mathbf{h}_{C_n} \in \Gamma_{K+|\mathcal{E}|}$ and $\mathbf{h}_{C_n;((X_1)^n,\ldots,(X_K)^n)} = \sum_{i=1}^k \mathbf{h}_{C_n;(X_i)^n}$ (the entry in $\mathbf{h}_{C_n}$ corresponding to the joint entropy of all sources is equal to the sum of entropies corresponding to the individual sources). Moreover,

$$\mathbf{h}_{C_n;\left(((X_s)^n, s\in S(o(e))), (\hat{g}^n_{e'}, d(e')=o(e))\right)} = \mathbf{h}_{C_n;\left(\hat{g}^n_e, ((X_s)^n, s\in S(o(e))), (\hat{g}^n_{e'}, d(e')=o(e))\right)} \tag{13}$$

and

$$\mathbf{h}_{C_n;\left(((X_s)^n, s\in D(t)), (\hat{g}^n_{e'}, d(e')=t)\right)} - \mathbf{h}_{C_n;\left(\hat{g}^n_{e'}, d(e')=t\right)} \leq n\delta_n,$$

where $\delta_n \to 0$. That is, the entries in $\mathbf{h}_{C_n}$ satisfy the functional constraints with equality and the decoding constraint with an inequality (but with a small error when normalized by n). As a result, for large enough n, $\frac{1}{n}\mathbf{h}_{C_n}$ is arbitrarily close to the polyhedral cone R_{out}. Since the rays of R_{out} are linearly representable, any point in R_{out} can be approximated by a linearly representable point (Lemma 2). Thus, for any $\zeta > 0$ and sufficiently large n we have

$$\|\frac{1}{n}\mathbf{h}_{C_n} - \mathbf{r}_\zeta\|_\infty < 2\zeta \tag{14}$$

where $\mathbf{r}_\zeta \in R_{out}$ is linearly representable.

Associated with $\mathbf{r}_\zeta$ are $K + |\mathcal{E}|$ random variables $\tilde{X}_1, \ldots, \tilde{X}_K, \{\tilde{X}_e\}_{e\in\mathcal{E}}$, which satisfy the following:

1) $H(\tilde{X}_i) = \frac{\log|F|}{\alpha}\mathbf{r}_{\zeta,(X_i)^n}$, where $\mathbf{r}_{\zeta,(X_i)^n}$ is the entry in $\mathbf{r}_\zeta$ corresponding to the source i.
2) The $\tilde{X}_1, \ldots, \tilde{X}_K$ are independent.
3) The $\tilde{X}_1, \ldots, \tilde{X}_K, \{\tilde{X}_e\}_{e\in\mathcal{E}}$ satisfy the functional and the decoding constraints with equality.

We first use the result in [9, Theorem 15.6] to show that from the random vector $\tilde{X}_1, \ldots, \tilde{X}_K, \{\tilde{X}_e\}_{e\in\mathcal{E}}$ one can construct a code with a small error probability. We then show that since $\tilde{X}_1, \ldots, \tilde{X}_K, \{\tilde{X}_e\}_{e\in\mathcal{E}}$ are uniform on their support and related to each other by solely a matrix multiplication (when they are dependent), the internal operations are linear.

The proof of [9, Theorem 15.6] is based on mapping the typical sequences of the original sources to those of the new random variables, $\tilde{X}_i$ in this case (with the appropriate block length), then, at each internal node, mapping the incoming sequences to the appropriate outgoing sequence using the fact that the functional constraint implies the existence of an appropriate function. Finally, when receiving the indices of the correct typical sequences at the decoders, mapping them back to the original inputs. Thus, we use the first step in the proof of [9, Theorem 15.6] to index the typical sequence of sources and map them to the typical $(\tilde{X}_i)^{n_o}$, for the appropriate block size n_0. Now, since auxiliary random variables $\{\tilde{X}_i\}$ are uniform, no compression is required, and the chosen sequence can be sent directly. At any internal node, since the auxiliary random variables $\tilde{X}_e$ corresponding to an edge e are functions of the $\{\tilde{X}_{e'}\}, e' : d(e') = o(e)$, they are linear functions of these variables, and the operation at the internal node is linear. At a terminal t, the values of $(\tilde{X}_i)^{n_0}, i \in D(t)$ are recovered without loss, and translated back to the original typical source vectors. ■

A consequence of Theorem 3 is that all points in $\mathcal{R}(\mathcal{V}, \mathcal{E}, \Pi_{i=1}^K p_i(x_i), S, D)$ can be represented as a convex combination of the entropies of m random vectors over a

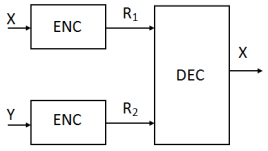

Fig. 1. Coded side information.

finite alphabet, independently of any block length n. Note that it suffices that the points $\{\mathbf{w}_i\}_{i=1}^m$ are entropic, and not necessarily linearly representable. We have

$$\mathcal{R}(\mathcal{V},\mathcal{E},\Pi_{i=1}^K p_i(x_i),S,D) \subseteq R_{out}|_{\{K+1,\ldots,K+|\mathcal{E}|\}}$$

and $R_{out} = \text{NonNeg}(\mathbf{w}_1,\ldots,\mathbf{w}_m)$. Since all $\{\mathbf{w}_i\}_{i=1}^m$ are linearly representable, by Theorem 2 there exist random vectors $(\{\hat{X}_e^1\}_{e\in\mathcal{E}})$ to $(\{\hat{X}_e^m\}_{e\in\mathcal{E}})$ such that any point $(R_e)_{e\in\mathcal{E}} \in \mathcal{R}(\mathcal{V},\mathcal{E},\Pi_{i=1}^K p_i(x_i),S,D)$ can be represented as a convex combination of the random variables.

Finally, note that while finding a linear representation for the entropy vectors might require an exhaustive search and is complex for large networks, it is straightforward for small networks and results in non-trivial statements regarding their rate region.

IV. Examples

In this section, we analyze two examples: the first one is a simple 3-node network for which it is easy to follow the suggested method. Note that if condition (7) is removed, the outer bound applies and it is easy to see what are the points spanning the polyhedral cone and why are they linearly representable. The second example is 4-node network which includes non-multicast demands. Moreover, this is a network for which the cut-set bound does not give tight results. Using the method suggested in this paper, we conclude that the outer bound for the network is linearly representable and Theorem 3 applies.

Example 3. Consider the coded side information problem in Figure 1. The receiver uses independent descriptions of X and Y to reconstruct X. Label the two inputs X and Y by A and B, respectively. Label X encoder output by C and the Y encoder output by D. In addition to all polymatroid axioms on A,B,C,D, we also have $w_A = w_{AC}$, $w_B = w_{BD}$ and $w_{CD} = w_{ACD}$. Hence, $R_{out} = \text{NonNeg}(\mathbf{w}_1,\ldots,\mathbf{w}_6)$, where the $\{\mathbf{w}_i\} \in \mathbb{R}^{15}$ are

$$\begin{aligned}
\mathbf{w}_1 &= (0,1,0,0,1,0,0,1,1,0,1,1,0,1,1)\\
\mathbf{w}_2 &= (0,1,0,1,1,0,1,1,1,1,1,1,1,1,1)\\
\mathbf{w}_3 &= (1,0,1,0,1,1,1,1,0,1,1,1,1,1,1)\\
\mathbf{w}_4 &= (1,1,1,0,1,1,1,1,1,1,1,1,1,1,1)\\
\mathbf{w}_5 &= (1,1,0,1,1,1,1,1,1,1,1,1,1,1,1)\\
\mathbf{w}_6 &= (1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)
\end{aligned}$$

These pseudo-entropy vectors are easily seen to be linearly representable, and the corresponding 6 random vectors are:

$$\begin{aligned}
V_1 &= (0,X_1,0,0) & V_4 &= (X_4,X_4,X_4,0)\\
V_2 &= (0,X_2,0,X_2) & V_5 &= (X_5,X_5,0,X_5)\\
V_3 &= (X_3,0,X_3,0) & V_6 &= (X_6,X_6,X_6,X_6)
\end{aligned}$$

where $\{X_i\}$ are random symmetric independent bits.

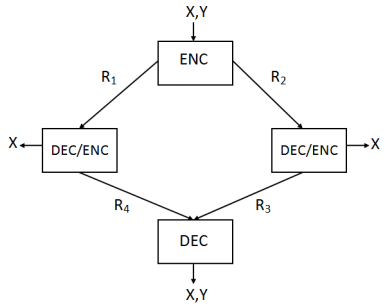

Fig. 2. A simple 4-node network which does not satisfy the cut-set bound.

Example 4. Consider the network given in Figure 2. This network is analyzed in [9, Section 15.1.1], where it is shown that the min-cut max-flow bound cannot be achieved for this network. Yet, it is not hard to check that the polyhedral cone of the outer bound is linearly representable, hence the results of Theorem 3 apply. This example demonstrates that the condition for linear representability of the outer bound do not coincide with the tightness of the cut-set bound.

References

[1] R. Ahlswede, N. Cai, S-Y. R. Li, and R. W. Yeung, "Network information flow," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[2] T. Ho, M. Médard, R. Koetter, D. R. Karger, M. Effros, J. Shi, and B. Leong, "A random linear network coding approach to multicast," *IEEE Trans. Inform. Theory*, vol. 52, no. 10, pp. 4413–4430, October 2006.

[3] R. Dougherty, C. Freling, and K. Zeger, "Insufficiency of linear coding in network information flow," *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2745–2759, August 2005.

[4] A. Cohen, S. Avestimehr, and M. Effros, "On networks with side information," in *Proc. of ISIT*, Seoul, Korea, June-July 2009.

[5] R. Yeung, S.-Y.R. Li, N. Cai, and Z. Zhang, "Network coding theory," *Foundation and Trends in Communications and Information Theory*, vol. 2, no. 4–5, pp. 241–381, 2005.

[6] T. H. Chan and A. Grant, "Dualities between entropy functions and network codes," *IEEE Trans. Inform. Theory*, vol. 54, no. 10, pp. 4470–4487, October 2008.

[7] L. Guille, T. H. Chan, and A. Grant, "The minimal set of ingleton inequalities," in *Proc. ISIT*, Toronto, Canada, July 2008.

[8] F. Matúš, "Piecewise linear conditional information inequality," *IEEE Trans. Inform. Theory*, vol. 52, no. 1, pp. 236–238, January 2006.

[9] R. W. Yeung, *A First Course in Information Theory*, Springer, 2002.

[10] J. G. Oxley, *Matroid Theory*, Oxford Univ. Press, Oxford, U.K., 1992.

[11] N. Y. Zolotykh, "Skeleton: Implementation of double description method," 2006, http://www.uic.nnov.ru/~zny/skeleton.

[12] D. Hammer, A. Romashchenko, A. Shen, and N. Vereshchagin, "Inequalities for shannon entropy and kolmogorov complexity," *Journal of Computer and System Sciences*, vol. 60, pp. 442–464, 2000.

[13] Z. Zhang and R. W. Yeung, "A non-shannon-type conditional inequality of information quantities," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1982–1986, November 1997.

[14] A. Cohen, M. Effros, S. Avestimehr, and R. Koetter, "Linearly representable entropy vectors and their relation to network coding solutions," *In preparation*, 2009.